

RANDOM MATRICES: SHARP CONCENTRATION OF EIGENVALUES

TERENCE TAO AND VAN VU

ABSTRACT. Let $W_n = \frac{1}{\sqrt{n}}M_n$ be a Wigner matrix whose entries have vanishing third moment, normalized so that the spectrum is concentrated in the interval $[-2, 2]$. We prove a concentration bound for $N_I = N_I(W_n)$, the number of eigenvalues of W_n in an interval I .

Our result shows that N_I decays exponentially with standard deviation at most $O(\log^{O(1)} n)$. This is best possible up to the constant exponent in the logarithmic term. As a corollary, the bulk eigenvalues are localized to an interval of width $O(\log^{O(1)} n/n)$; again, this is optimal up to the exponent. These results strengthen recent results of Erdős, Yau and Yin (under the extra assumption of vanishing third moment).

Our proof is relatively simple and relies on the Lindeberg replacement argument.

1. INTRODUCTION

The purpose of this paper is to sharpen the existing bounds on the eigenvalue counting function $N_I = N_I(W_n)$ of a (normalized) Wigner matrix $W_n = \frac{1}{\sqrt{n}}M_n$, and related quantities such as the Stieltjes transform $s_{W_n}(z)$ and individual eigenvalues $\lambda_i(W_n)$. Let us first state the Wigner random matrix model which we will use.

Definition 1 (Wigner matrices). Let $n \geq 1$ be an integer (which we view as a parameter going off to infinity; in particular, n is understood to be large enough that quantities such as $\log \log n$ are well-defined and positive). An $n \times n$ *Wigner matrix* M_n is defined to be a random Hermitian $n \times n$ matrix $M_n = (\xi_{ij})_{1 \leq i, j \leq n}$, in which the ξ_{ij} for $1 \leq i \leq j \leq n$ are jointly independent with $\xi_{ji} = \overline{\xi_{ij}}$ (in particular, the ξ_{ii} are real-valued). For $1 \leq i < j \leq n$, we require that the ξ_{ij} have mean zero and variance one, while for $1 \leq i = j \leq n$ we require that the ξ_{ij} (which are necessarily real) have mean zero and variance σ^2 for some $\sigma^2 > 0$ independent of i, j, n . For simplicity, we will also assume that for each $1 \leq i < j \leq n$, the real and imaginary parts $\operatorname{Re}\xi_{ij}$, $\operatorname{Im}\xi_{ij}$ are independent. We refer to the distributions $\operatorname{Re}\xi_{ij}$, $\operatorname{Im}\xi_{ij}$ for $1 \leq i < j \leq n$ and ξ_{ii} for $1 \leq i \leq n$ as the *atom distributions* of M_n , and view them as fixed while n goes off to infinity.

1991 *Mathematics Subject Classification.* 15A52.

T. Tao is supported by a grant from the MacArthur Foundation, and by NSF grant DMS-0649473.

V. Vu is supported by research grants DMS-0901216 and AFOSAR-FA-9550-09-1-0167.

We say that the Wigner matrix ensemble *obeys Condition C0* if we have the exponential decay condition

$$(1) \quad \mathbf{P}(|\xi_{ij}| \geq t^C) \leq e^{-t}$$

for all $1 \leq i, j \leq n$ and $t \geq C'$, and some constants C, C' (independent of i, j, n).

Two Wigner matrices $M_n = (\xi_{ij})_{1 \leq i, j \leq n}$ and $M'_n = (\xi'_{ij})_{1 \leq i, j \leq n}$ are said to have *matching moments to order m* for some $m \geq 0$ if one has

$$(2) \quad \mathbf{E} \operatorname{Re}(\xi_{ij})^k \operatorname{Im}(\xi_{ij})^l = \mathbf{E} \operatorname{Re}(\xi'_{ij})^k \operatorname{Im}(\xi'_{ij})^l$$

for all $1 \leq i, j \leq n$ and all natural numbers $k, l \geq 0$ with $k + l \leq m$. As we are assuming the real and imaginary parts to be independent, this condition simplifies to the conditions

$$(3) \quad \mathbf{E} \operatorname{Re}(\xi_{ij})^k = \mathbf{E} \operatorname{Re}(\xi'_{ij})^k; \quad \mathbf{E} \operatorname{Im}(\xi_{ij})^k = \mathbf{E} \operatorname{Im}(\xi'_{ij})^k$$

for all $1 \leq i, j \leq n$ and all $0 \leq k \leq m$. If we only require (2) or (3) to hold in the off-diagonal case $i \neq j$ (resp. in the diagonal case $i = j$), we say that M_n and M'_n *match moments to order m off the diagonal* (resp. *on the diagonal*).

We observe four basic examples of Wigner matrices:

- In the *Gaussian Unitary Ensemble* (GUE), $\xi_{ij} \equiv N(0, 1)_{\mathbb{C}}$ is the standard complex gaussian random variable for $1 \leq i < j \leq n$, $\xi_{ii} \equiv N(0, 1)_{\mathbb{R}}$ is the standard real gaussian random variable for $1 \leq i \leq n$, and $\sigma^2 = 1$.
- In the *Gaussian Orthogonal Ensemble* (GOE) $\xi_{ij} \equiv N(0, 1)_{\mathbb{R}}$ is the standard real gaussian random variable for $1 \leq i < j \leq n$, $\xi_{ii} \equiv N(0, 2)_{\mathbb{R}}$ is a slightly rescaled real gaussian random variable for $1 \leq i \leq n$, and $\sigma^2 = 2$.
- In the *symmetric Bernoulli ensemble*, ξ_{ij} equals $+1$ with probability $1/2$ and -1 with probability $1/2$ for all $1 \leq i, j \leq n$, and $\sigma^2 = 1$.
- In the *complex Hermitian Bernoulli ensemble*, $\operatorname{Re} \xi_{ij}, \operatorname{Im} \xi_{ij}$ for $1 \leq i < j \leq n$ and ξ_{ii} for $1 \leq i \leq n$ all equal $+1$ with probability $1/2$ and -1 with probability $1/2$, and $\sigma^2 = 1$.

Remark 2. Note that we do not require the off-diagonal ξ_{ij} , $1 \leq i < j \leq n$ (or the diagonal ξ_{ii} , $1 \leq i \leq n$) to be identically distributed. This lack of an identical distribution hypothesis will be convenient when we apply the Lindeberg exchange strategy [27], in which one Wigner matrix is compared to another one by exchanging the entries of the former matrix with the latter one¹ at a time. As such, the intermediate stages of this exchange process need not have identically distributed entries, even if the initial and final matrices do.

The hypothesis of independence of real and imaginary parts is imposed purely to simplify the exposition, and can easily be removed at the cost of some more complicated notation; in particular, the simpler moment matching condition (3) must be replaced by the more complicated condition (2). See Remark 23.

¹More precisely, we exchange the diagonal entries one at a time, and the off-diagonal entries two at a time, in order to preserve the Hermitian property throughout.

In this paper, we will mostly deal with the (coarse-scale) normalization $W_n := \frac{1}{\sqrt{n}}M_n$ of M_n of the Wigner matrix, and more specifically with the eigenvalue counting function

$$N_I = N_I(W_n) := |\{1 \leq i \leq n : \lambda_i(W_n) \in I\}|$$

of this matrix for various intervals $I \subset \mathbb{R}$, where $\lambda_1(W_n) \leq \dots \leq \lambda_n(W_n)$ denote the (necessarily) real eigenvalues of the (Hermitian) matrix W_n .

The well-known *Wigner semicircle law* describes the bulk behavior of the counting function N_I of a Wigner matrix in terms of the *semicircular distribution* $\rho_{sc}(x) dx$, where $\rho_{sc} : \mathbb{R} \rightarrow \mathbb{R}$ is the function

$$\rho_{sc}(x) := \frac{1}{2\pi}(4 - x^2)_+^{1/2}.$$

Theorem 3 (Semicircular law). *Let M_n be a Wigner Hermitian matrix obeying Condition **C0**. Then for any fixed interval I (independent of n), one has*

$$\lim_{n \rightarrow \infty} \frac{1}{n} N_I(W_n) = \int_I \rho_{sc}(y) dy$$

in the sense of probability.

See for instance [4] for a proof of this theorem and for historical background. Condition **C0** can be omitted from this law, but we retain the hypothesis as it will be needed for the subsequent results discussed below.

If we use $o(x)$ to denote a quantity that goes to zero as $n \rightarrow \infty$ after dividing by x , we can reformulate Theorem 3 as the assertion that the asymptotic

$$(4) \quad N_I(W_n) = n \int_I \rho_{sc}(y) dy + o(n)$$

holds with probability $1 - o(1)$ for each fixed I .

One can also phrase the semicircular law in terms of the individual eigenvalues $\lambda_i(W_n)$. If for each $1 \leq i \leq n$ we define the *classical location* γ_i of the normalised i^{th} eigenvalue by the formula

$$(5) \quad \int_{-\infty}^{\gamma_i} \rho_{sc}(x) dx = \frac{i}{n}.$$

then the Wigner semicircular law (combined with an almost sure bound of $(2 + o(1))\sqrt{n}$ for the operator norm of M_n , due to Bai and Yin [5]) is equivalent to the assertion that one has

$$(6) \quad \lambda_i(W_n) = \gamma_i + o(1)$$

for any given $1 \leq i \leq n$, with probability $1 - o(1)$.

In this paper we investigate sharper versions of the semicircular law (known in the literature as *local semicircular laws*), which improve upon the error terms and failure probabilities in (4) and (6), and in which the interval I is now allowed to depend on n .

We first discuss the case of the Gaussian Unitary Ensemble (GUE), which is the most well-understood case, as the joint distribution of the eigenvalues is given by a determinantal point process. Because of this, it is known that for any interval I , the random variable $N_I(W_n)$ in the GUE case obeys a law of the form

$$(7) \quad N_I(W_n) \equiv \sum_{i=1}^{\infty} \eta_i$$

where the $\eta_i = \eta_{i,n,I}$ are jointly independent indicator random variables (i.e. they take values in $\{0, 1\}$); see e.g. [3, Corollary 4.2.24]. The mean and variance of $N_I(W_n)$ can also be computed in the GUE case with a high degree of accuracy:

Theorem 4 (Mean and variance for GUE). *Let M_n be drawn from GUE, let $W_n := \frac{1}{\sqrt{n}}M_n$, and let $I = [-\infty, x]$ for some real number x (which may depend on n). Let $\varepsilon > 0$ be independent of n .*

(i) (Bulk case) *If $x \in [-2 + \varepsilon, 2 - \varepsilon]$, then*

$$\mathbf{E}N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O\left(\frac{\log n}{n}\right).$$

(ii) (Edge case) *If $x \in [-2, 2]$, then*

$$\mathbf{E}N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O(1).$$

(iii) (Variance bound) *If one has $x \in [-2, 2 - \varepsilon]$ and $n^{2/3}(2+x) \rightarrow \infty$ as $n \rightarrow \infty$, one has*

$$\mathbf{Var}N_I(W_n) = \left(\frac{1}{2\pi^2} + o(1)\right) \log(n(2+x)^{3/2}).$$

In particular, one has $\mathbf{Var}N_I(W_n) = O(\log n)$ in this regime.

Here of course we use $X = O(Y)$, $X \ll Y$ or $Y \gg X$ to denote the estimate $|X| \leq CY$ for some quantity C independent of n . We will also use c to denote various small positive constants $c > 0$ independent of n (but possibly depending on the constants in Condition **C0**).

Proof. See [19, Lemmas 2.1, 2.2, 2.3]. Note that the normalization conventions in [19] differ by a factor of $\sqrt{2}$ from the ones used here. Also, the asymptotic in the statement of [19, Lemma 2.2] is only accurate (with the $O(1)$ error term) for t sufficiently close to 1, and more precisely for $t = 1 - O(n^{-2/5})$ (or, in our notation, $x = -2 + O(n^{-2/5})$), as it implicitly relies on the approximation $n \int_I \rho_{\text{sc}}(y) dy = \frac{2}{3\pi}(2+x)^{3/2} + O(1)$ (as written in our notation), which is only valid in this regime. \square

By combining these estimates with a well-known inequality of Bennett [6], we obtain a concentration estimate for $N_I(W_n)$ in the GUE case:

Corollary 5 (Concentration for GUE). *Let M_n be drawn from GUE, let $W_n := \frac{1}{\sqrt{n}}M_n$, and let I be an interval. Then one has*

$$\mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll \exp(-cT)$$

for all $T \gg \log n$.

Proof. By the triangle inequality we may take $I = [-\infty, x]$ for some real number x . As ρ_{sc} is supported on $[-2, 2]$ and has total mass 1, we see (using the trivial bounds $0 \leq N_I(W_n) \leq n$ and $N_I(W_n) \leq N_J(W_n)$ whenever $I \subset J$) that without loss of generality we may assume $x \in [-2, 2]$. By (7) and Theorem 4, $N_I(W_n)$ is then the sum of independent indicator functions whose mean μ and variance σ^2 are given by

$$\mu = n \int_I \rho_{\text{sc}}(y) dy + O(1)$$

and $\sigma^2 = O(\log n)$ respectively. Bennett's inequality (see [6], or [21, p.29]) then asserts that

$$\mathbf{P}(|N_I(W_n) - \mu| \geq t) \leq 2 \exp(-\sigma^2 \phi(\frac{t}{\sigma^2}))$$

where $\phi(x) := (1+x)\log(1+x) - x$. Since $\phi(x) \gg x$ when $x \gg 1$, the claim follows. \square

Let us say that an event holds with *overwhelming probability* if it occurs with probability $1 - O(n^{-A})$ for each fixed A . From the above corollary we see in particular that in the GUE case, one has

$$N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O(\log^{1+o(1)} n)$$

with overwhelming probability for each fixed I , and an easy union bound argument (ranging over all intervals I in, say, $[-3, 3]$ whose endpoints are a multiple of n^{-100} (say)) then shows that this is also true uniformly in I as well.

Remark 6. By using a general result of Costin and Lebowitz [7], one can also obtain a central limit theorem for $N_I(W_n)$ as long as I is not too small; see [19]. Such results have also been recently been extended to more general Wigner matrices in [8]. However, such theorems will not be the focus of the current paper.

Now we turn from the GUE case to more general Wigner ensembles. There has been much interest in recent years in obtaining concentration results for $N_I(W_n)$ (and for closely related objects, such as the Stieltjes transform $s_{W_n}(z) := \frac{1}{n} \text{trace}(W_n - z)^{-1}$ of W_n) for short intervals I , due to the applicability of such results to establishing various universality properties of such matrices; see [11, 12, 13, 30, 31, 14, 16, 17]. The previous best result in this direction was by Erdős, Yau, and Yin [17] (see also [9] for a variant):

Theorem 7. [17] *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Then, for any interval I , one has*

$$(8) \quad \mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll \exp(-cT^c)$$

for all $T \geq \log^{A \log \log n} n$, and some constant $A > 0$.

Proof. See [17, Theorem 2.2]. □

One can reformulate (8) equivalently as the assertion that

$$\mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll \exp(\log^{O(\log \log n)} n) \exp(-cT^c)$$

for all $T > 0$.

In particular, this theorem asserts that with overwhelming probability one has

$$N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O(\log^{O(\log \log n)} n)$$

for all intervals I . The proof of the above theorem is somewhat lengthy, requiring a delicate analysis of the self-consistent equation of the Stieltjes transform of W_n .

Comparing this result with the previous results for the GUE case, we see that there is a loss of a double logarithm $\log \log n$ in the exponent. The first main result of this paper² is to remove this double logarithmic loss, at least under an additional vanishing moment assumption:

Theorem 8 (First main theorem). *Let M_n be a Wigner matrix obeying Condition C0, and let $W_n := \frac{1}{\sqrt{n}} M_n$. Assume that M_n matches moments with GUE to third order off the diagonal (i.e. $\text{Re} \xi_{ij}, \text{Im} \xi_{ij}$ have variance $1/2$ and third moment zero). Then, for any interval I , one has*

$$\mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll n^{O(1)} \exp(-cT^c)$$

for any $T > 0$.

This estimate is phrased for any T , but the bound only becomes non-trivial when $T \gg \log^C n$ for some sufficiently large C . In that regime, we see that this result removes the double-logarithmic factor from Theorem 7. In particular, this theorem implies that with overwhelming probability one has

$$N_I(W_n) = n \int_I \rho_{\text{sc}}(y) dy + O(\log^{O(1)} n)$$

for all intervals I ; in particular, for any I , $N_I(W_n)$ has variance $O(\log^{O(1)} n)$.

Remark 9. As we are assuming $\text{Re}(\xi_{ij})$ and $\text{Im}(\xi_{ij})$ to be independent, the moment matching condition simplifies to the constraints that $\mathbf{E} \text{Re}(\xi_{ij})^2 = \mathbf{E} \text{Im}(\xi_{ij})^2 = \frac{1}{2}$ and $\mathbf{E} \text{Re}(\xi_{ij})^3 = \mathbf{E} \text{Im}(\xi_{ij})^3 = 0$. However, it is possible to extend this theorem to the case when the real and imaginary parts of ξ_{ij} are not independent; see Remark 23.

²We would like to thank M. Ledoux for a private conversation that led to this question.

Remark 10. The constant c in the bound in Theorem 8 is quite decent in several cases. For instance, if the atom variables of M_n are Bernoulli or have sub-gaussian tail, then we can set $c = 2/5 - o(1)$ by optimizing our arguments (details omitted). If we assume 4 matching moments rather than 3, then we can set $c = 1$ (see Remark 26), matching the bound in Corollary 5. It is an interesting question to determine the best value of c . The value of c in [16] is implicit and rather small.

We prove Theorem 8 in Sections 2-4. Our argument is different from that in [17] in that it only uses a relatively crude analysis of the self-consistent equation to obtain some preliminary bounds on the Stieltjes transform and on N_I (which were also essentially implicit in previous literature). Instead, the bulk of the argument relies on using the Lindeberg swapping strategy to deduce concentration of $N_I(W_n)$ in the non-GUE case from the concentration results in the GUE case provided by Corollary 5. In order to keep the error terms in this swapping under control, three matching moments³ are needed.

Very roughly speaking, the main idea of the argument is to show that high moments such as

$$\mathbf{E}|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy|^k$$

are quite stable (in a multiplicative sense) if one swaps (the real or imaginary part of) one of the entries of W_n (and its adjoint) with another random variable that matches the moments of the original entry to third order. For technical reasons, however, we do not quite manipulate $N_I(W_n)$ directly, but instead work with a proxy for this quantity, namely a certain integral of the Stieltjes transform of W_n . As observed in [16], the Lindeberg swapping argument is quite simple to implement at the level of the Stieltjes transform (due to the simplicity of the resolvent identities, when compared against the rather complicated Taylor expansions of individual eigenvalues used in [30]).

The result in Theorem 8 is well suited for controlling eigenvalues in the bulk of the spectrum, but is not sufficient by itself to control eigenvalues at the edge, and in particular the largest eigenvalue $\lambda_1(W_n)$ and the smallest eigenvalue $\lambda_n(W_n)$. However, it is known that these eigenvalues are highly concentrated around $+2$ and -2 respectively. In the GUE case, we have the following concentration result of Ledoux [24] and Aubrun [1] (see also [25], [26] for further discussion and refinements):

Theorem 11 (Concentration for GUE). [24, 1] *Let M_n be drawn from GUE, let $W_n := \frac{1}{\sqrt{n}}M_n$. Then one has*

$$\mathbf{P}(n^{2/3}(\lambda_1(W_n) - 2) \geq T) \ll \exp(-cT^{3/2})$$

for all $T > 0$. By symmetry, we also have

$$\mathbf{P}(n^{2/3}(-\lambda_n(W_n) - 2) \geq T) \ll \exp(-cT^{3/2}).$$

³Compare with the “four moment theorem” from [30]. We need one less moment here because we are working at “mesoscopic” scales (in which the number of eigenvalues involved is much larger than 1) rather than at “microscopic” scales. However, in Theorem 14 below, only one eigenvalue is involved, making the problem microscopic enough to require four moments instead of three.

Remark 12. As is well known, the random variable $n^{2/3}(\lambda_1(W_n) - 2)$ in fact converges in distribution to the Tracy-Widom law [34]. However, we will not focus on this law here. The exponent $3/2$ on the right-hand side cannot be improved (indeed, it matches the decay rate of the Tracy-Widom law); see [1] for further discussion.

This result was partially extended to the Wigner case in [17]:

Theorem 13. [17] *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Then one has*

$$(9) \quad \mathbf{P}(n^{2/3}(\lambda_1(W_n) - 2) \geq T) \ll \exp(-cT^c)$$

for all $T \geq \log^{A \log \log n} n$, for some $A > 0$ independent of n . By symmetry, one also has

$$\mathbf{P}(n^{2/3}(-\lambda_n(W_n) - 2) \geq T) \ll \exp(-cT^c).$$

Proof. See [17, Theorem 2.1]. □

As before, we can reformulate (9) equivalently as the assertion that

$$\mathbf{P}(n^{2/3}(\lambda_1(W_n) - 2) \geq T) \ll \exp(\log^{O(\log \log n)} n) \exp(-cT^c)$$

for all $T > 0$.

Our second main result is to remove the double logarithm from Theorem 13, at the cost of requiring matching GUE to fourth order rather than to third order:

Theorem 14 (Second main theorem). *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Assume that M_n matches moments with GUE to fourth order off the diagonal and second order on the diagonal (i.e. $\sigma^2 = 1$). Then one has*

$$\mathbf{P}(n^{2/3}(\lambda_1(W_n) - 2) \geq T) \ll n^{O(1)} \exp(-cT^c)$$

for any $T > 0$. By symmetry, one then also has

$$\mathbf{P}(n^{2/3}(-\lambda_n(W_n) - 2) \geq T) \ll n^{O(1)} \exp(-cT^c)$$

We will derive Theorem 14 from Theorem 11 in Section 5 using the same techniques used to derive Theorem 8 from Corollary 5.

By combining Theorem 8 and Theorem 14 one can “solve” for individual eigenvalues $\lambda_i(W_n)$ to obtain an appropriate concentration (localization) result:

Corollary 15 (Concentration of eigenvalues). *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Assume that M_n matches moments with GUE to three order off the diagonal and second order on the diagonal. Then for any $\min(i, n - i + 1) \geq \log^{A \log \log n} n$ for some sufficiently large A , we have*

$$\mathbf{P}(n^{2/3} \min(i, n - i + 1)^{1/3} |\lambda_i(W_n) - \gamma_i| \geq T) \ll n^{O(1)} \exp(-cT^c)$$

for any $T > 0$, where the classical location $\gamma_i \in [-2, 2]$ is defined by the formula

$$\int_{-2}^{\gamma_i} \rho_{\text{sc}}(y) dy = \frac{i}{n}.$$

If we assume for moment, then the estimate holds for all $1 \leq i \leq n$.

The first part of this corollary significantly improves [30, Theorem 29]. (As a matter of fact, the original proof of this theorem has a gap in it; see [33, Appendix A] for a further discussion.)

Proof. By Theorems 8, 14 and the union bound, we see that outside of an event of probability $n^{O(1)} \exp(-cT^c)$, we have

$$N_I = n \int_I \rho_{\text{sc}}(y) dy + O(T)$$

for all intervals I , as well as the bounds

$$-2 - O(n^{-2/3}T) \leq \lambda_n(W_n) \leq \lambda_1(W_n) \leq 2 + O(n^{-2/3}T).$$

Some elementary estimation of the semicircular density ρ_{sc} and its integrals $\int_I \rho_{\text{sc}}(y) dy$ (cf. [17, §5]) then gives

$$\lambda_i(W_n) = \gamma_i + O(n^{-2/3} \min(i, n-i+1)^{-1/3}T)$$

for all $1 \leq i \leq n$. The claim follows (possibly after adjusting T by a multiplicative factor). \square

Remark 16. The results in this paper also hold if one replaces the GUE ensemble by the GOE ensemble. To do this, one needs to replace Theorem 4 and Theorem 11 by their GOE counterparts. The GOE version of Theorem 4 was established by O'Rourke [29]. The GOE version of Theorem 11 can be deduced from the GUE version (possibly at the expense of worsening the $3/2$ exponent) by using the connection between the GOE and GUE point processes observed by Forrester and Rains [18]; we omit the details. In principle, one might be able to use other ensembles (such as the gaussian divisible matrices [22]) to match moments with, which would allow one to remove the moment conditions almost entirely. We will not pursue these matters here.

We thank Michel Ledoux and Atti Knowles for supplying some relevant references.

2. REDUCTION TO THE STIELTJES TRANSFORM

We now begin the proof of Theorem 8. The first step is to replace the counting function $N_I = N_I(W_n)$ with the *Stieltjes transform* s_{W_n} , defined by the formula

$$(10) \quad s_{W_n}(z) := \frac{1}{n} \text{trace}(W_n - z)^{-1} = \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i(W_n) - z}$$

for any complex number z with positive imaginary part. We can express this Stieltjes transform as a Riemann-Stieltjes integral

$$(11) \quad s_{W_n}(z) = \frac{1}{n} \int_{\mathbb{R}} \frac{1}{x-z} dN_{(-\infty, x)}.$$

which gives a clear connection between the Stieltjes transform and the counting function. Using the heuristic $dN_{(-\infty, x)} \approx n\rho_{\text{sc}}(x) dx$ from (4), we thus expect to have $s_{W_n} \approx s_{\text{sc}}$, where

$$s_{\text{sc}}(z) := \int_{\mathbb{R}} \frac{1}{x-z} \rho_{\text{sc}}(x) dx.$$

As is well known, s_{sc} can be evaluated explicitly via contour integration as

$$(12) \quad s_{\text{sc}}(z) = \frac{1}{2}(-z + \sqrt{z^2 - 4}),$$

where $\sqrt{z^2 - 4}$ is the branch of the square root that is asymptotic to z at infinity. In particular, s_{sc} exactly obeys the *self-consistent equation*

$$(13) \quad s_{\text{sc}}(z) = -\frac{1}{s_{\text{sc}}(z) + z}$$

In the case of GUE, we may easily formalize this heuristic with the assistance of Corollary 5:

Proposition 17 (Concentration for GUE). *Let M_n be drawn from GUE, and $W_n := \frac{1}{\sqrt{n}}M_n$. Then for any $T > 0$ and any complex number $z = E + \sqrt{-1}\eta$ with $\eta > 0$, one has*

$$\mathbf{P}(|s_{W_n}(z) - s_{\text{sc}}(z)| \geq \frac{T}{n\eta}) \ll n^{O(1)} \exp(-cT).$$

Proof. We may assume that $T \gg \log n$, as the claim is trivial otherwise. Let $T_1 \gg \log n$ be chosen later. From Corollary 5 and the union bound, we see that with probability $1 - O(n^{O(1)} \exp(-cT_1))$, one has

$$|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \ll T_1$$

for all intervals I in $[-3, 3]$ whose endpoints are multiples of n^{-100} , and hence for all intervals I . In particular,

$$N_{(-\infty, x)} = n \int_{-\infty}^x \rho_{\text{sc}}(y) dy + O(T_1)$$

for all x . On the other hand, from (11) and integration by parts, one has

$$s_{W_n}(z) = \frac{1}{n} \int_{\mathbb{R}} \frac{1}{(x-z)^2} N_{(-\infty, x)} dx.$$

A similar integration by parts gives

$$s_{\text{sc}}(z) = \int_{\mathbb{R}} \frac{1}{(x-z)^2} \left(\int_{-\infty}^x \rho_{\text{sc}}(y) dy \right) dx,$$

and thus by the triangle inequality

$$s_{W_n}(z) = s_{\text{sc}}(z) + O\left(\frac{1}{n} \int_{\mathbb{R}} \frac{1}{|x-z|^2} T_1 dx\right).$$

The error term on the right-hand side evaluates to $O(\frac{T}{n\eta})$. The claim then follows by choosing T_1 to be a small multiple of T . \square

We will use this proposition to obtain a similar concentration result for Wigner matrices:

Theorem 18 (Concentration for Wigner). *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Assume that M_n matches moments with GUE to third order off the diagonal. Then for any $T > 0$ and any complex number $z = E + \sqrt{-1}\eta$ with $E \in [-3, 3]$ and $0 < \eta \ll n^{100}$, one has*

$$\mathbf{P}(|s_{W_n}(z) - s_{\text{sc}}(z)| \geq \frac{T}{n\eta}) \ll n^{O(1)}(\exp(-cT^c) + \exp(-c(n\eta)^c)).$$

We prove this theorem in later sections. Let us assume it for now, and use it to establish Theorem 8. Let M_n, W_n, T, K be as in the above theorem. By the triangle inequality, we may take $I = (-\infty, E)$ for some real number E ; from the support of ρ_{sc} , we may assume that $E \in [-2, 2]$. We may also take $T \gg \log^{100} n$ (say), as the claim is trivial otherwise.

Let $T_1 \gg T/\log n \gg \log^{99} n$ be a quantity to be chosen later, and set $\eta_0 := T_1/n$. Applying Theorem 18 and the union bound, we see that outside of an event of probability at most

$$(14) \quad n^{O(1)} \exp(cT_1^{-c}),$$

one has

$$(15) \quad |s_{W_n}(E + \sqrt{-1}\eta) - s_{\text{sc}}(E + \sqrt{-1}\eta)| \ll \frac{T_1}{n\eta}$$

for all η between η_0 and n^{100} which is a multiple of n^{-200} . On the other hand, in this range one easily verifies that s_{W_n} and s_{sc} are Lipschitz with Lipschitz norm at most $O(n^{10})$ (say). As a consequence, we conclude (after conditioning outside of the above exceptional event) that (15) holds for *all* η between η_0 and n^{100} .

By conditioning on another event of probability at most (14), we may assume that all entries of M_n are of size at most $O(n)$ (say). Among other things, this implies that all eigenvalues $\lambda_i(W_n)$ are (very crudely) of size at most $O(n^{20})$.

Since $\eta \geq \eta_0 = T_1/n$, we conclude from (15) and (12) that

$$|s_{W_n}(E + \sqrt{-1}\eta)| \ll 1.$$

On the other hand, from (10) one has

$$\text{Im} s_{W_n}(E + \sqrt{-1}\eta) = \frac{1}{n} \sum_{i=1}^n \frac{\eta}{|\lambda_i(W_n) - E|^2 + \eta^2}$$

and in particular

$$\text{Im} s_{W_n}(E + \sqrt{-1}\eta) \gg \frac{1}{n\eta} N_{[E-\eta, E+\eta]}.$$

We conclude that

$$(16) \quad N_{[E-\eta, E+\eta]} \ll n\eta$$

for all $\eta \geq \eta_0$ (note that this claim is trivial for $\eta \geq n^{100}$). (One could also have used Proposition 30 at this juncture.)

Next, if we integrate (15) and use the triangle inequality we observe that

$$(17) \quad \operatorname{Re} \int_{\eta_0}^{n^{100}} s_{W_n}(E + \sqrt{-1}\eta) d\eta = \operatorname{Re} \int_{\eta_0}^{n^{100}} s_{\text{sc}}(E + \sqrt{-1}\eta) d\eta + O\left(\frac{T_1 \log n}{n}\right).$$

Let us now evaluate the left-hand side. From the definition of the Stieltjes transform, we may rewrite it as

$$\frac{1}{n} \sum_{i=1}^n \operatorname{Arg}(E + \sqrt{-1}\eta_0 - \lambda_i(W_n)) - \operatorname{Arg}(E + \sqrt{-1}n^{100} - \lambda_i(W_n)),$$

where Arg is the standard branch of the argument on the upper half-plane.

Since $E \in [-2, 2]$ and $\lambda_i(W_n) = O(n^{20})$, we have

$$\operatorname{Arg}(E + \sqrt{-1}n^{100} - \lambda_i(W_n)) = \frac{\pi}{2} + O(n^{-50})$$

(say). Also, from elementary trigonometry one has

$$\operatorname{Arg}(E + \sqrt{-1}\eta_0 - \lambda_i(W_n)) = \pi 1_{\lambda_i(W_n) \geq E} + O\left(\frac{\eta_0}{|\lambda_i(W_n) - E| + \eta_0}\right).$$

We may therefore write the left-hand side of (17) as

$$\frac{\pi}{2} - \frac{1}{n} \pi N_{(-\infty, E)} + O\left(\frac{1}{n} \sum_{i=1}^n \frac{\eta_0}{|\lambda_i(W_n) - E| + \eta_0}\right) + O(n^{-50}).$$

On the other hand, from (16) and dyadic decomposition (recalling that $\lambda_i(W_n) = O(n^{20})$) one has

$$\frac{1}{n} \sum_{i=1}^n \frac{\eta_0}{|\lambda_i(W_n) - E| + \eta_0} = O(\eta_0 \log n)$$

and thus

$$\operatorname{Re} \int_{\eta_0}^{n^{100}} s_{W_n}(E + \sqrt{-1}\eta) d\eta = \frac{\pi}{2} - \frac{1}{n} \pi N_{(-\infty, E)} + O\left(\frac{T_1 \log n}{n}\right).$$

A similar argument gives

$$\operatorname{Re} \int_{\eta_0}^{n^{100}} s_{\text{sc}}(E + \sqrt{-1}\eta) d\eta = \frac{\pi}{2} - \frac{1}{n} \pi \int_{-\infty}^E \rho_{\text{sc}}(y) dy + O\left(\frac{T_1 \log n}{n}\right).$$

From (17) we thus conclude that

$$N_{(-\infty, E)} = \int_{-\infty}^E \rho_{\text{sc}}(y) dy + O(T_1 \log n).$$

Choosing T_1 to be a small multiple of $T/\log n$ (and bounding T_1^c from below by $T^{c'} - O(\log n)$ for some sufficiently small $c' > 0$), we obtain Theorem 8 as desired.

It remains to deduce Theorem 18 from Proposition 17. This will be the objective of the next few sections.

3. THE MOMENT METHOD, AND THE LINDBERG STRATEGY

Given a matrix $W_n = \frac{1}{\sqrt{n}}M_n$ and a complex number $z = E + \sqrt{-1}\eta$, define the quantity $A(W_n) = A(W_n, z)$ by the formula

$$A(W_n) := n\eta(s_{W_n}(z) - s_{sc}(z)).$$

This quantity describes the normalised deviation of the Stieltjes transform of W_n from the semicircular law at z . In this notation, Proposition 17 becomes the assertion that

$$(18) \quad \mathbf{P}(|A(W_n)| \geq T) \ll n^{O(1)} \exp(-cT)$$

whenever $T > 0$, $E \in \mathbb{R}$, and $\eta > 0$, when M_n is drawn from GUE. Similarly, Theorem 18 becomes the assertion that

$$(19) \quad \mathbf{P}(|A(W_n)| \geq T) \ll n^{O(1)}(\exp(-cT^c) + \exp(-c(n\eta)^c))$$

whenever $T > 0$, $E \in [-3, 3]$, and $0 < \eta \ll n^{100}$, when M_n is drawn from a Wigner matrix obeying Condition **C0**, and with $\text{Re}\xi_{ij}$ and $\text{Im}\xi_{ij}$ having variance $1/2$ and third moment zero for $1 \leq i < j \leq n$.

To deduce (19) from (18) we will use the moment method combined with the Lindeberg exchange strategy; more specifically, we will show that a high moment $\mathbf{E}A(W_n)^k$ for some large even number k (which one should think of, in practice, as comparable to T) is stable under the operation of replacing (the real or imaginary part of) one entry of M_n (and its transpose) with another entry with a number of matching moments. The Lindeberg exchange strategy is by now a standard tool in establishing universality properties for Wigner matrices [30], [16]; the main novelty here⁴ is the application of that strategy to a high moment $\mathbf{E}A(W_n)^k$ (as opposed to a quantity such as $\mathbf{E}G(A(W_n))$ for some smooth test function G).

Let us now make the strategy more precise. Let us call two Wigner matrices M_n, M'_n *real-adjacent*, or *adjacent* for short, if their respective atom variables ξ_{ij}, ξ'_{ij} are equal except for a single choice of $(i, j) = (a, b)$ and its transpose $(i, j) = (b, a)$, and such that ξ_{ab}, ξ'_{ab} either have identical real parts, or identical imaginary parts. Thus, a Wigner matrix M'_n adjacent to M_n is formed by changing the real or imaginary part of a single entry of M_n and its adjoint, leaving the other components of M_n unchanged. The main technical step is then to establish the following proposition.

Proposition 19 (Stability of moments). *Let M_n, M'_n be two adjacent Wigner matrices obeying Condition **C0**, whose moments match to order m for some fixed*

⁴Very recently [23], a similar application of the Lindeberg exchange strategy to a high moment of a spectral statistic was used to establish some related concentration results. We thank Antti Knowles for bringing this preprint to our attention.

$m = O(1)$. Let $z = E + \sqrt{-1}\eta$ for some $E \in [-3, 3]$ and $0 < \eta \ll n^{100}$, and set $W_n := \frac{1}{\sqrt{n}}M_n$ and $W'_n := \frac{1}{\sqrt{n}}M'_n$. Then for any even integer $k \geq \log n$, one has

$$\mathbf{E}A(W_n)^k \leq \left(1 + O\left(\frac{1}{n^{(m+1)/2}}\right)\right) \mathbf{E}A(W'_n)^k + O(k)^k + O(n^{O(k)} \exp(-(n\eta)^c)). \quad (20)$$

Let us assume this proposition for now and establish Theorem 18. Let $n, M_n, W_n, E, \eta, z, T$ be as in that theorem. We may assume that $T \geq \log^{C_0} n$ (say) for some sufficiently large absolute constant C_0 , as the claim is trivial otherwise; we may also assume that $T \leq \eta n$, since the claim follows from existing local semicircle laws (in particular, Corollary 32). In particular, we may now assume that $T \leq n^{O(1)}$ and $\eta \geq \log^{C_0} n/n$. Our task is now to show that

$$\mathbf{P}(|A(W_n)| \geq T) \ll n^{O(1)} \exp(-cT^c). \quad (21)$$

On the other hand, if M'_n is drawn from GUE and $W'_n := \frac{1}{\sqrt{n}}M'_n$, then from Proposition 17 one has

$$\mathbf{P}(|A(W'_n)| \geq T) \ll n^{O(1)} \exp(-cT)$$

for all $T > 0$. In particular, for any $k \geq \log n$, one has

$$\begin{aligned} \mathbf{E}|A(W'_n)|^k &= \int_0^\infty \mathbf{P}(|A(W'_n)| \geq T) kT^{k-1} dT \\ &\ll kn^{O(1)} \int_0^\infty e^{-cT} T^{k-1} dT \\ &\ll O(1)^k n^{O(1)} k! \\ &\ll O(k)^k \end{aligned} \quad (22)$$

We can replace M'_n with M_n in a sequence of n^2 exchanges from one Wigner matrix to a real-adjacent one; $n^2 - n$ of these exchanges arise by swapping the real or imaginary part of an off-diagonal entry ξ_{ij} of M'_n (and its transpose ξ_{ji}) with the corresponding component of M_n , and n of these exchanges arise by swapping a diagonal entry ξ_{ii} of M'_n with the corresponding entry of M_n . We perform these exchanges in an arbitrary order. By hypothesis, for the $n^2 - n$ off-diagonal exchanges one has matching moments to order $m = 3$, while for the diagonal exchanges one has matching moments to order $m = 1$. Applying Proposition 19 $n^2 - n$ times with $m = 3$ and n times with $m = 1$ and concatenating, we conclude that for any $k \geq \log n$ one has

$$\mathbf{E}A(W_n)^k \leq O(1)\mathbf{E}A(W'_n)^k + O(n^2)O(k)^k + O(n^{O(k)} \exp(-(n\eta)^c)).$$

In particular, from (22) one has

$$\mathbf{E}A(W_n)^k \ll O(k)^k + O(n^{O(k)} \exp(-(n\eta)^c))$$

and hence by Markov's inequality

$$\mathbf{P}(|A(W_n)| \geq T) \ll \left(\frac{O(k)}{T}\right)^k + T^{-k} n^{O(k)} \exp(-(n\eta)^c).$$

If we set k to be the largest even integer less than T^{c_0} for some absolute constant c_0 , and if C_0 is sufficiently large depending on c_0 , we obtain (21) as desired, thanks to the assumptions $\log^{C_0} n \leq T \leq n\eta$.

Remark 20. An inspection of the above argument reveals that we in fact have the slight refinement

$$\mathbf{P}(|A(W_n)| \geq T) \ll n^{O(1)} \exp(-cT)$$

in the regime $T \leq (n\eta)^c$, since in this regime we may take k to be a small multiple of T (rounded off to the nearest even integer, of course). Unfortunately, this refinement does not appear to immediately offer any significant improvement to the conclusion of Theorem 8.

It remains to establish Proposition 19. This will be achieved in the next section.

4. STABILITY OF HIGH MOMENTS

We now prove Proposition 19. We introduce a definition:

Definition 21 (Elementary matrix). An *elementary matrix* is a matrix which has one of the following forms

$$(23) \quad V = e_a e_a^* e_b e_b^* + e_b e_a^* \sqrt{-1} e_a e_b^* - \sqrt{-1} e_b e_a^*$$

with $1 \leq a, b \leq n$ distinct, where e_1, \dots, e_n is the standard basis of \mathbb{C}^n .

As M_n, M'_n are real-adjacent, one can write

$$M_n = M_n^0 + \xi V; \quad M'_n = M_n^0 + \xi' V$$

for some elementary matrix V , some random matrix M_n^0 , and some real random variables ξ, ξ' independent of M_n^0 that match moments to m^{th} order and obey the exponential decay condition

$$(24) \quad \mathbf{P}(|\xi| \geq t^C), \mathbf{P}(|\xi'| \geq t^C) \leq e^{-t}$$

for all $t \geq C'$ and some $C, C' > 0$.

We now recall some (deterministic) resolvent stability results concerning matrices of the form $M_n^0 + tV$. Define the matrix norm $\|R\|_{(\infty,1)}$ of a $n \times n$ matrix $R = (R_{ij})_{1 \leq i,j \leq n}$ by the formula

$$\|R\|_{(\infty,1)} := \sup_{1 \leq i,j \leq n} |R_{ij}|.$$

Proposition 22 (Stability of resolvent). *Let M_n^0 be a Hermitian matrix, let V be an elementary matrix, and let t be a real number. Let $z := E + \sqrt{-1}\eta$ be a complex number with $\eta > 0$. Write*

$$R_t := (M_n^0 + tV - z)^{-1}$$

and suppose that

$$|t| \|R_0\|_{(\infty,1)} = o(\sqrt{n}).$$

Then

$$\|R_t\|_{(\infty,1)} = (1 + o(1)) \|R_0\|_{(\infty,1)}.$$

Furthermore, if we set $s_t := \frac{1}{n} \text{trace } R_t$, then we have the Taylor expansion

$$s_t = s_0 + \sum_{j=1}^m n^{-j/2} c_j t^j + O(n^{-(m+1)/2} |t|^{m+1} \|R_0\|_{(\infty,1)}^{m+1} \min(\|R_0\|_{(\infty,1)}, \frac{1}{n\eta}))$$

for any fixed nonnegative $m = O(1)$, where the coefficients c_j are independent of t and obey the bounds

$$(25) \quad |c_j| \ll \|R_0\|_{(\infty,1)}^j \min(\|R_0\|_{(\infty,1)}, \frac{1}{n\eta}).$$

for all $1 \leq j \leq m$.

Proof. See [32, Lemma 12] and [32, Proposition 13]. \square

Our objective is to establish (20). From Corollary 33 we see that

$$\|R_\xi\|_{(\infty,1)} = O(1)$$

with probability $1 - O(n^{O(1)} \exp(-(n\eta)^c))$, while from (24) we certainly have $\xi = o(\sqrt{n})$ with $1 - O(n^{O(1)} \exp(-(n\eta)^c))$. Hence by Proposition 22 (reversing the roles of R_0 and R_ξ) we have

$$(26) \quad \|R_0\|_{(\infty,1)} = O(1)$$

with probability $1 - O(n^{O(1)} \exp(-(n\eta)^c))$. Using the crude bound $A(W_n) = O(n^{O(1)})$, we may thus condition M_n^0 to be fixed and obeying (26), since the contribution of the event where (26) fails to $\mathbf{E}A(W_n)^k$ is $O(n^{O(k)} \exp(-(n\eta)^c))$.

By Proposition 22, we thus see that whenever $\xi = o(\sqrt{n})$, one has

$$(27) \quad A(W_n) = A_0 + \sum_{j=1}^m a_j (\xi/\sqrt{n})^j + O((|\xi|/\sqrt{n})^{m+1})$$

where the coefficients A_0, a_j are deterministic (and in particular independent of ξ, ξ'), and a_j obeys the bound $a_j = O(1)$.

Suppose first that $|A_0| \leq k$. Then one has

$$|A(W_n)| \ll k$$

whenever $\xi = o(\sqrt{n})$, which gives a net contribution of $O(k)^k$ to $\mathbf{E}|A(W_n)|^k$; meanwhile, from (24), the case when $\eta \gg \sqrt{n}$ contributes at most $O(n^{O(k)} \exp(-(n\eta)^c))$. Thus we may assume that $|A_0| > k$. Thus we have

$$A(W_n) = A_0(1 + \frac{1}{k}(\sum_{j=1}^m b_j (\xi/\sqrt{n})^j + O((\xi/\sqrt{n})^{m+1})))$$

for some deterministic coefficients $b_1, \dots, b_m = O(1)$, and assuming that $\eta = o(\sqrt{n})$. Raising this to the k^{th} power (after using Taylor's theorem with remainder to expand $(1 + \frac{1}{k}x)^k$ to m^{th} order in the regime $x = o(1)$), we conclude that

$$A(W_n)^k = A_0^k(1 + \sum_{j=1}^m d_j (\xi/\sqrt{n})^j + O((|\xi|/\sqrt{n})^{m+1}))$$

for some deterministic coefficients $d_1, \dots, d_m = O(1)$ (which are allowed to depend on k), whenever $\xi = o(\sqrt{n})$. Taking (conditional) expectations in η (using (24) and the trivial bound $A(W_n) = O(n^{O(1)})$ to handle the tail event when $|\eta| \gg \sqrt{n}$) we conclude that

$$\mathbf{E}(A(W_n)^k | M_n^0) = A_0^k (1 + \sum_{j=1}^m d_j n^{-j/2} \mathbf{E} \xi^j + O(n^{-(m+1)/2})) + O(n^{O(k)} \exp(-(n\eta)^c)).$$

and thus

$$\mathbf{E} A(W_n)^k = \mathbf{E}(A_0^k (1 + \sum_{j=1}^m d_j n^{-j/2} \mathbf{E} \xi^j + O(n^{-(m+1)/2}))) + O(n^{O(k)} \exp(-(n\eta)^c)) + O(k)^k.$$

Similarly we have

$$\mathbf{E} A(W_n)^k = \mathbf{E}(A_0^k (1 + \sum_{j=1}^m d_j n^{-j/2} \mathbf{E}(\xi')^j + O(n^{-(m+1)/2}))) + O(n^{O(k)} \exp(-(n\eta)^c)) + O(k)^k.$$

Since ξ and ξ' match to order k , we obtain the claim. This concludes the proof of Proposition 19 and hence Theorem 8.

Remark 23. It is possible to adapt the above arguments to the case when $\operatorname{Re} \xi_{ij}$ and $\operatorname{Im} \xi_{ij}$ are not assumed to be independent. The main new difficulty is that instead of swapping the real and imaginary parts of a single entry ξ_{ab} of M_n (and its transpose ξ_{ba}) separately, one has to swap them together. This requires one to consider perturbations of the form

$$M_n = M_n^0 + \xi_1 V_1 + \xi_2 V_2$$

where V_1, V_2 are two distinct elementary random variables, and ξ_1, ξ_2 are real random variables that are not necessarily independent and obeying the exponential decay hypothesis (24). However, it is possible to extend Proposition 22 without much difficulty to the case of two-parameter perturbations and perform a similar argument to that given above. We omit the details.

5. EXTREME EIGENVALUES

We now prove Theorem 14, by combining the arguments in previous sections with some ideas from [17] (and in particular, demonstrating a concentration of $\operatorname{Im} s_{W_n}(E + \sqrt{-1}\eta)$ that is better than $1/n\eta$ for some energy $E > 2$). By symmetry, it suffices to prove the bound for $\lambda_1(W_n)$. We may of course assume that n is large.

By standard large deviation estimates, one has

$$\mathbf{P}(\lambda_1(W_n) \geq E) \ll \exp(-cn^c \log E)$$

for any $E \geq 3$; see⁵ [16, Lemma 7.2]. This already deals with the case when $n^{2/3} \leq T \leq n^{100}$ (say), and the case $T > n^{100}$ can be handled by crudely bounding $\lambda_1(W_n)$ by, say, the Frobenius norm of W_n and using Condition **C0**. Thus we may restrict attention to the regime $T \leq n^{2/3}$, and show that

$$\mathbf{P}(2 + n^{-2/3}T \leq \lambda_1(W_n) \leq 3) \ll n^{O(1)} \exp(-cT^c).$$

⁵One could also use the earlier estimates in [28] or [2]; see also [3] for more discussion.

We may assume that $T \geq \log^{C_0} n$ for some suitably large absolute constant C_0 , as the claim is trivial otherwise.

Suppose that $\lambda_1(W_n)$ was in the interval $[2 + n^{-2/3}T, 3]$. Set $\eta := n^{-2/3}$, and let $B(W_n)$ denote the quantity

$$B(W_n) := n\eta \text{Im}s_{W_n}(E + \sqrt{-1}\eta).$$

From the identity

$$(28) \quad B(W_n) = \sum_{i=1}^n \frac{\eta^2}{|\lambda_i(W_n) - E|^2 + \eta^2}$$

we conclude in particular that

$$B(W_n) \geq \frac{1}{10}$$

where E is the closest multiple of $n^{-2/3}$ in $[2 + n^{-2/3}T, 3]$ to $\lambda_1(W_n)$. Thus, by the union bound, it will suffice to show that

$$(29) \quad \mathbf{P}(B(W_n) \geq \frac{1}{10}) \ll n^{O(1)} \exp(-cT^c)$$

for any fixed $E \in [2 + n^{-2/3}T, 3]$.

Let M'_n be drawn from GUE, and set $W'_n := \frac{1}{\sqrt{n}}M'_n$. By Theorem 11, we have

$$\lambda_1(W'_n) \leq 2 + n^{-2/3}T/2$$

outside of an event of probability $O(\exp(-cT^{3/2}))$. Also, from Corollary 5 and the union bound we see that outside of an event of probability $O(n^{O(1)} \exp(-cT^c))$, one has

$$N_I(W'_n) \leq \int_I \rho_{sc}(y) dy + O(T^{0.1})$$

(say) for all intervals I . In particular, we see that outside of an event of probability $O(n^{O(1)} \exp(-cT^c))$, one has

$$N_{[E - n^{-2/3}T/2, E + n^{-2/3}T/2]} = 0$$

and

$$N_{[E - 2^k n^{-2/3}T, E + 2^k n^{-2/3}T]} \ll 2^{3k/2} T^{3/2}$$

for all $k \geq 1$. From this, (28), and dyadic decomposition one easily establishes that

$$B(W'_n) \ll \frac{1}{T^{1/2}}$$

outside of an event of probability $O(n^{O(1)} \exp(-cT^c))$.

Let $\log n \leq k \leq n^{0.01}$ be an integer to be chosen later. Since we may trivially bound $\text{Im}s_{W_n}(E + \sqrt{-1}\eta)$ by $n^{O(1)}$, we conclude that

$$(30) \quad \mathbf{E}B(W'_n)^k \ll O\left(\frac{1}{T}\right)^{k/2} + n^{O(k)} \exp(-cT^c).$$

We claim the following stability result for $\mathbf{E}B(W_n)^k$, analogous to Proposition 19:

Proposition 24 (Stability of moments). *Let M_n, M'_n be two adjacent Wigner matrices obeying Condition **C0**, whose moments match to order m for some fixed $m = O(1)$. Set $W_n := \frac{1}{\sqrt{n}}M_n$ and $W'_n := \frac{1}{\sqrt{n}}M'_n$. Then for any integer $\log n \leq k \leq n^{0.1}$, one has*

$$(31) \quad \mathbf{E}B(W_n)^k \leq (1 + O((k/\sqrt{n})^{m+1}))\mathbf{E}B(W'_n)^k + O(100^{-k}) + O(n^{O(k)} \exp(-cT^c)).$$

Applying this proposition $n^2 - n$ times with $m = 4$ and n times with $m = 2$ we conclude that

$$\mathbf{E}B(W_n)^k \ll (1 + O(k^5/n^{5/2}))^{n^2-n} (1 + O(k^3/n^{3/2}))^n (\mathbf{E}B(W'_n)^k + O(n^{O(1)} 100^{-k}) + O(n^{O(k)} \exp(-cT^c)))$$

and thus (using (30) and the hypothesis $k \leq n^{0.01}$)

$$\mathbf{E}B(W_n)^k \ll n^{O(1)} 100^{-k} + n^{O(k)} \exp(-cT^c).$$

The desired claim (29) then follows from Markov's inequality by taking $k = T^{c_0}$ for some sufficiently small $c_0 > 0$ (and assuming C_0 sufficiently large depending on $c_0 > 0$).

It remains to establish Proposition 24. As in the previous section, we write

$$M_n = M_n^0 + \xi V; \quad M'_n = M_n^0 + \xi' V$$

for some elementary matrix V , some random matrix M_n^0 , and some real random variables ξ, ξ' independent of M_n^0 that match moments to m^{th} order and obey the exponential decay condition (24). Arguing exactly as before, we may condition M_n^0 to be a deterministic matrix for which

$$\|R_0\|_{(\infty,1)} = O(1).$$

Using Proposition 22 as before, we see that

$$B(W_n) = B_0 + \sum_{j=1}^m a_j (\xi/\sqrt{n})^j + O((|\xi|/\sqrt{n})^{m+1})$$

for some deterministic coefficients B_0 and $a_j = O(1)$, whenever $\xi = o(\sqrt{n})$.

Suppose first that $|B_0| \leq 1/200$. Then one has $|B(W_n)| \leq 1/100$ whenever $\xi = o(\sqrt{n})$, and so this case contributes $O(100^{-k}) + O(n^{O(k)} \exp(-cn^c))$ to (31), which is acceptable. Thus we may restrict attention to the case when $|B_0| > 1/200$. Then we may write

$$B(W_n) = B_0(1 + \sum_{j=1}^m b_j (\xi/\sqrt{n})^j + O((|\xi|/\sqrt{n})^{m+1}))$$

whenever $\xi = o(\sqrt{n})$, where the $b_j = O(1)$ are deterministic coefficients.

Suppose now that $\xi = O(n^{0.3})$. Since $k \leq n^{0.01}$, we may perform a Taylor expansion of $(1+x)^k$ to order m for $x = O(n^{-0.2})$ and conclude that

$$B(W_n)^k = B_0^k(1 + \sum_{j=1}^m c_j (k\xi/\sqrt{n})^j + O((k|\xi|/\sqrt{n})^{m+1}))$$

in this regime, where the $c_j = O(1)$ are deterministic coefficients (which are allowed to depend in k). Taking expectations as in the preceding section, and using (24) to handle those ξ with $|\xi| \geq n^{0.3}$, we conclude that

$$\mathbf{E}B(W_n)^k = \mathbf{E}(B_0^k(1 + \sum_{j=1}^m c_j k^j n^{-j/2} \mathbf{E}\xi^j + O((k/\sqrt{n})^{m+1}))) + O(n^{O(k)} \exp(-(n\eta)^c)) + O(100^{-k}),$$

and similarly for $\mathbf{E}B(W'_n)^k$; and the claim follows from the matching moments hypothesis.

Remark 25. As in Remark 23, it is possible to extend these arguments to the case when $\text{Re}(\xi_{ij})$ and $\text{Im}(\xi_{ij})$ are not independent; we leave the details to the interested reader.

Remark 26. Note that when one has four matching moments rather than three, the error terms are more favorable by a factor of \sqrt{n} , giving some additional room to vary the parameters of the argument by small powers of n . Because of this, it is possible to modify the proof of Theorem 18 to conclude in this case that

$$\mathbf{P}(|A(W_n)| \geq T) \ll n^{O(1)} \exp(-cT)$$

in the regime $0 < T \leq n^c$ for a sufficiently small c . This is achieved by arguing as in this section, except that one allows the resolvent $\|R_0\|_{(\infty,1)}$ to be as large as $O(n^c)$ rather than $O(1)$ in order to keep the failure probability bounded by $O(n^{O(1)} \exp(-n^c))$ rather than $O(n^{O(1)} \exp(-(n\eta)^c))$. We omit the details. As a consequence, we can sharpen the conclusion of Theorem 8 to

$$\mathbf{P}(|N_I(W_n) - n \int_I \rho_{\text{sc}}(y) dy| \geq T) \ll n^{O(1)} \exp(-cT)$$

when $0 < T \leq n^c$ and M_n matches moments with GUE to fourth order off the diagonal and second order on the diagonal.

APPENDIX A. LOCAL SEMICIRCLE LAW

In this appendix we establish some preliminary local semicircle law estimates, following the treatment in [16] and [30]. As the methods used here are now standard, and the results very close to those in [16] and [30], we shall be somewhat brief in our treatment.

We first recall a concentration estimate of Hanson and Wright [20].

Proposition 27 (Concentration of quadratic forms). *Let $X \in \mathbb{C}^n$ be a vector of independent random variables ξ_1, \dots, ξ_n of mean zero and variance σ^2 , obeying the uniform subexponential decay bound*

$$\mathbf{P}(|\xi_i| \geq t^C \sigma) \leq e^{-t}$$

for all $t \geq C'$ and $1 \leq i \leq n$, and some $C, C' > 0$ independent of n . Let A be an $n \times n$ matrix. Then for any $T > 0$, one has

$$\mathbf{P}(|X^*AX - \sigma^2 \text{trace } A| \geq T\sigma^2(\text{trace}(A^*A))^{1/2}) \ll \exp(-cT^c).$$

Thus

$$X^*AX = \sigma^2(\text{trace } A + O(T \text{trace}(A^*A))^{1/2})$$

outside of an event of probability $O(\exp(-cT^c))$.

Proof. See [16, Lemma B.1]. (Note that a factor of σ is missing from the statement of the exponential decay hypothesis in the lemma as stated in [16], which is needed in order to reduce to the $\sigma = 1$ case.) \square

Corollary 28 (Distance between a random vector and a subspace). *Let X and σ be as in Proposition 27, and let V be a d -dimensional complex subspace of \mathbb{C}^n . Let π_V be the orthogonal projection to V . Then one has*

$$0.9d\sigma^2 \leq \|\pi_V(X)\|^2 \leq 1.1d\sigma^2$$

outside of an event of probability $O(\exp(-cd^c))$.

Proof. Apply the preceding proposition with $A := \pi_V$ (so $\text{trace } A = \text{trace } A^*A = d$) and $T := d^{1/2}/10$. \square

Remark 29. We can also use Talagrand's inequality as in [30], combining with a truncation argument (to bound each entries by some properly chosen quantity K). In the case when the atom variables have very fast decay (such as sub-gaussian) or bounded (such as Bernoulli), this calculation will actually lead to a decent bound on the value of c in Theorem 8.

We can now establish a crude upper bound on the counting function N_I of a Wigner matrix.

Proposition 30 (Crude upper bound). *Let M_n be a Wigner matrix obeying Condition C0, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Then for any interval I , one has*

$$N_I(W_n) = O(n|I|)$$

outside of an event of probability $O(n^{O(1)} \exp(-c(n|I|)^c))$.

Proof. Fix I , which we write as $I = [E - \eta, E + \eta]$. Suppose that

$$(32) \quad N_I(W_n) \geq Cn\eta$$

for some sufficiently large absolute constant C to be chosen later. We will show that this leads to a contradiction outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$.

From the identity

$$\text{Im} s_{W_n}(E + \sqrt{-1}\eta) = \frac{1}{n} \sum_{i=1}^n \frac{\eta}{|\lambda_i(W_n) - E|^2 + \eta^2}$$

and (32), we see that

$$\text{Im} s_{W_n}(E + \sqrt{-1}\eta) \gg C.$$

On the other hand, we can write the Stieltjes transform s_{W_n} in terms of the coefficients R_{ij} of the resolvent as

$$s_{W_n}(E + \sqrt{-1}\eta) = \frac{1}{n} \sum_{i=1}^n R_{ii}(E + \sqrt{-1}\eta).$$

Thus, by the pigeonhole principle, we have

$$\operatorname{Im} R_{ii}(E + \sqrt{-1}\eta) \gg C$$

for some $1 \leq i \leq n$. By symmetry (and conceding a factor of n in the failure probability estimates) we may take $i = n$.

Now, a standard Schur complement computation (see e.g. [30, Lemma 42]) shows that

$$(33) \quad R(z)_{nn} = \frac{1}{\frac{1}{\sqrt{n}}\xi_{nn} - z - X^*R^{(n)}(z)X}$$

where $R^{(n)}(z) = (W_n^{(n)} - z)^{-1}$ is the resolvent corresponding to the $(n-1) \times (n-1)$ matrix $W_n^{(n)}$ formed by removing the n^{th} row and column from W_n , ξ_{nn} is the bottom right entry of M_n , and X is the rightmost column of W_n (after removing the bottom entry $\frac{1}{\sqrt{n}}\xi_{nn}$). In particular, using the trivial bound $|\operatorname{Im} \frac{1}{z}| \leq \frac{1}{|\operatorname{Im} z|}$, we conclude that

$$\operatorname{Im} R_{nn}(E + \sqrt{-1}\eta) \leq \frac{1}{\eta + \operatorname{Im} X^*R^{(n)}(E + \sqrt{-1}\eta)X} \leq \frac{1}{\operatorname{Im} X^*R^{(n)}(z)X}$$

and thus

$$\operatorname{Im} X^*R^{(n)}(E + \sqrt{-1}\eta)X \ll C^{-1}.$$

Now, by the Cauchy interlacing law, $W_n^{(n)}$ has $\gg Cn\eta$ consecutive eigenvalues in I . There are $O(n^2)$ possibilities for the starting and ending index of these eigenvalues. If we let V be the space spanned by the corresponding eigenvectors, then $\dim(V) \gg Cn\eta$, and from the spectral theorem we see that

$$\operatorname{Im} X^*R^{(n)}(E + \sqrt{-1}\eta)X \gg \|\pi_V(X)\|^2/\eta$$

and thus

$$\|\pi_V(X)\|^2 \ll \frac{1}{C}\eta.$$

On the other hand, from (28) we see that

$$\|\pi_V(X)\|^2 \gg C\eta$$

outside of an event of probability $O(\exp(-c(n\eta)^c))$. If C is sufficiently large, the claim follows. \square

This gives rise to a self-consistent equation:

Proposition 31 (Self-consistent equation). *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Then for any $z = E + \sqrt{-1}\eta$ with $E = O(1)$ and $0 < \eta \ll n^{100}$, and all $1 \leq i \leq n$, one has*

$$R(z)_{ii} = -\frac{1}{s_{W_n}(z) + z + o(1)}$$

outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$. In particular, by the union bound, we have

$$(34) \quad s_{W_n}(z) = -\frac{1}{s_{W_n}(z) + z + o(1)}$$

outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$.

Proof. We can assume that $n\eta \geq \log^{100} n$ (say), as the claim is trivial otherwise. By symmetry, it will suffice to establish

$$R(z)_{nn} = -\frac{1}{s_{W_n}(z) + z + o(1)}$$

outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$. By (33), this statement is equivalent to

$$X^* R^{(n)}(z) X - \frac{1}{\sqrt{n}} \xi_{nn} = s_{W_n}(z) + o(1).$$

By Condition **C0**, one has $\frac{1}{\sqrt{n}} \xi_{nn} = o(1)$ outside of an event of probability $O(\exp(-cn^c))$, which is certainly acceptable; so our task is now to show that

$$(35) \quad X^* R^{(n)}(z) X = s_{W_n}(z) + o(1)$$

outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$.

From the Cauchy interlacing law (cf. [30, §5.2]) we know that

$$\frac{1}{n} \text{trace } R^{(n)}(z) = s_{W_n}(z) + o(1).$$

Also,

$$(36) \quad \text{trace } R^{(n)}(z)^* R^{(n)}(z) = \sum_{i=1}^{n-1} \frac{1}{|\lambda_i(W_n^{(n)}) - E|^2 + \eta^2}.$$

By Proposition 30 and the union bound, we may assume outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$, one has

$$N_I(W_n) \ll n|I|$$

for all intervals I of width at least η centered at E . By interlacing, we may also conclude

$$N_I(W_n^{(n)}) \ll n|I|$$

for such intervals. Inserting this bound into (36), we conclude that

$$(37) \quad \text{trace } R^{(n)}(z)^* R^{(n)}(z) \ll \frac{n}{\eta}.$$

If we then apply Proposition 27 with $T := (n\eta)^{1/4}$ (say), using the hypothesis that $n\eta \geq \log^{100} n$ (so that $1/(n\eta)^c = o(1)$ for any $c > 0$) we conclude (35) outside of an event of order $O(n^{O(1)} \exp(-c(n\eta)^c))$, as required. \square

We can combine this proposition with a standard stability analysis of the self-consistent equation (34) to conclude a crude version of the local semicircle law:

Corollary 32 (Local semicircle law). *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}} M_n$. Then for any $z = E + \sqrt{-1}\eta$ with $E = O(1)$ and $0 < \eta \ll n^{100}$, and all $1 \leq i \leq n$, one has*

$$(38) \quad s_{W_n}(z) = s_{\text{sc}}(z) + o(1)$$

and

$$(39) \quad R(z)_{ii} = s_{\text{sc}}(z) + o(1)$$

outside of an event with probability $O(n^{O(1)} \exp(-c(n\eta)^c))$.

We note that this corollary is essentially [17, Theorem 3.1]; in the statement of the result in [17] the additional constraint $\eta \geq \log^{C \log \log n} / n$ for some constant C is imposed, but this constraint is not actually used in the proof, at least if one is not concerned with obtaining the best possible bounds for the $o(1)$ error terms. For the convenience of the reader, we sketch the proof of this corollary below.

Proof. As before we may assume that $\eta \geq \log^{100} n / n$; we may also assume that n is large. By Proposition 31, we may assume that (34) holds.

Let us first dispose of the case when η is large, say $\eta \geq 100$. In this case, the imaginary part of $s_{W_n}(z) + z + o(1)$ is at least $100 - o(1)$, and hence by (34) one has $|s_{W_n}(z)| \leq 1/100 + o(1)$; inserting this back into (34) (and using (12)) one obtains $|s_{W_n}(z) - s_{sc}(z)| \leq 1/10$ (say). One can then deduce (38) from (34) (and (13)) by a routine application of the contraction mapping theorem.

Henceforth we assume that $\eta < 100$, so that $z = O(1)$. Then equation (34) already implies that $s_{W_n}(z) = O(1)$, since (34) cannot hold if $|s_{W_n}(z)|$ is too large. We may thus multiply out the denominator and conclude that

$$s_{W_n}(z)^2 + z s_{W_n}(z) + 1 = o(1).$$

Since the two solutions to the quadratic equation $s^2 + z s + 1 = 0$ are $s = s_{sc}(z)$ and $s = -z - s_{sc}(z)$, we conclude that

$$s_{W_n}(z) = s_{sc}(z) + o(1) \text{ or } s_{W_n}(z) = -z - s_{sc}(z) + o(1)$$

outside of an event with probability $O(n^{O(1)} \exp(-c(n\eta)^c))$.

We apply this fact with z replaced by an arbitrary complex numbers ζ with $\operatorname{Re}(\zeta) = O(1)$ and $\eta \leq \operatorname{Im}(\zeta) \ll 1$, and whose real and imaginary parts are multiples of n^{-100} (say). By the union bound, the probability of the failure event is still $O(n^{O(1)} \exp(-c(n\eta)^c))$. We may then remove the latter hypotheses using the fact that s_{W_n} and s_{sc} have Lipschitz constant $O(n)$ in this region, and conclude that outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$, one has

$$(40) \quad s_{W_n}(\zeta) = s_{sc}(\zeta) + o(1) \text{ or } s_{W_n}(\zeta) = -\zeta - s_{sc}(\zeta) + o(1)$$

for *all* ζ with $\operatorname{Re}(\zeta) = O(1)$ and $\eta \leq \operatorname{Im}(\zeta) \ll 1$. On the other hand, if one has $\operatorname{Im}(\zeta) \geq c$ for some absolute constant $c > 0$, then the second possibility in (40) cannot occur for n large enough, because $s_{W_n}(\zeta)$ necessarily has positive imaginary part. A continuity argument then shows that the first option in (40) holds for all ζ in the indicated region⁶. This gives (38). Among other things, this shows that $|s_{W_n}(z) + z| \gg 1$ (thanks to (13)), and then from (13) and the second part of Proposition 31 we obtain (39). \square

For our applications, we will also need bounds on the coefficient norm

$$\|R(z)\|_{(\infty,1)} := \sup_{1 \leq i,j \leq n} |R(z)_{ij}|$$

⁶When ζ approaches the edges ± 2 of the spectrum, thus $\zeta = \pm 2 + o(1)$, the two options in (40) begin to overlap, but in that regime one can deduce the first option from the second (with a slightly worse $o(1)$ error) and so the claim made in the text is still valid.

of the resolvent.

Corollary 33 (Resolvent bound). *Let M_n be a Wigner matrix obeying Condition **C0**, and let $W_n := \frac{1}{\sqrt{n}}M_n$. Then for any $z = E + \sqrt{-1}\eta$ with $E = O(1)$ and $0 < \eta \ll n^{100}$, one has*

$$(41) \quad \|R(z)\|_{(\infty,1)} = O(1)$$

outside of an event with probability $O(n^{O(1)} \exp(-c(n\eta)^c))$.

Proof. Again, we may assume $\eta > \log^{100} n/n$. By the union bound, it suffices to show for each $1 \leq i, j \leq n$ that

$$|R(z)_{ij}| = O(1)$$

outside of an event with probability $O(n^{O(1)} \exp(-c(n\eta)^c))$. In the diagonal case $i = j$, this follows directly from (39), so suppose that $i \neq j$. In this case, we may use the Schur complement identity

$$R(z)_{ij} = -R(z)_{ii}R^{(i)}(z)_{jj}K_{ij}^{(ij)}$$

where $R^{(i)}(z)$ is the resolvent associated to the $(n-1) \times (n-1)$ matrix $W_n^{(i)}$ formed by removing the i^{th} row and column from W_n , and $K_{ij}^{(ij)}$ is the quantity

$$K_{ij}^{(ij)} = \frac{1}{\sqrt{n}}\zeta_{ij} - X_i^*(W_n^{(ij)} - z)^{-1}X_j,$$

ζ_{ij} is the ij coefficient of W_n , $W_n^{(ij)}$ is the $(n-2) \times (n-2)$ matrix formed by removing the i^{th} and j^{th} rows and columns from W_n , and $X_i, X_j \in \mathbb{C}^{n-2}$ are the i^{th} and j^{th} columns of W_n , after removing the i^{th} and j^{th} rows. See [16, Lemma 4.2] for a proof of this identity. From (39) applied to both the original Wigner matrix W_n and the minor $W_n^{(i)}$ (which is essentially also a Wigner matrix, up to an easily manageable multiplicative factor of $\frac{\sqrt{n-1}}{\sqrt{n}}$) we see that $R(z)_{ii} = O(1)$ and $R^{(i)}(z)_{jj} = O(1)$ outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$, so it suffices to obtain the bound $K_{ij}^{(ij)} = O(1)$ outside of a similar event. But from Condition **C0**, one has $\frac{1}{\sqrt{n}}\zeta_{ij} = O(1)$ outside of an event of probability $O(\exp(-n^c))$, which is certainly acceptable, so it suffices to show that

$$X_i^*(W_n^{(ij)} - z)^{-1}X_j = O(1)$$

outside of an event of probability $O(n^{O(1)} \exp(-c(n\eta)^c))$. But by Proposition 27 (viewing the $(n-2) \times (n-2)$ matrix $(W_n^{(ij)} - z)^{-1}$ as the upper-right block of a nilpotent $2(n-2) \times 2(n-2)$ matrix, and concatenating X_i and X_j together), one has

$$X_i^*(W_n^{(ij)} - z)^{-1}X_j = O\left(\frac{1}{n}T(\text{trace}(((W_n^{(ij)} - z)^{-1})^*(W_n^{(ij)} - z)^{-1}))^{1/2}\right)$$

outside of an event of probability $O(\exp(-cT^c))$, for any $T > 0$. But by repeating the derivation of (37), one has

$$\text{trace}(((W_n^{(ij)} - z)^{-1})^*(W_n^{(ij)} - z)^{-1}) = O\left(\frac{n}{\eta}\right).$$

If one then sets $T = O(\sqrt{n\eta})$, one obtains the claim. \square

We remark that the above argument in fact shows that we may improve the bound $R(z)_{ij} = O(1)$ to $R(z)_{ij} = O(\frac{1}{(n\eta)^{1/2-\delta}})$ for any fixed $\delta > 0$; compare with [17, Theorem 3.1]. However, we will not need this improvement here.

REFERENCES

- [1] G. Aubrun, A sharp small deviation inequality for the largest eigenvalue of a random matrix, *Séminaire de Probabilités XXXVIII*, 320-337, Lecture Notes in Math., 1857, Springer, Berlin, 2005.
- [2] N. Alon, M. Krivelevich, V.H. Vu, On the concentration of eigenvalues of random symmetric matrices, *Israel J. Math.* **131** (2002) 259-267.
- [3] G. Anderson, A. Guionnet and O. Zeitouni, An introduction to random matrices, Cambridge Studies in Advanced Mathematics, 118. Cambridge University Press, Cambridge, 2010.
- [4] Z. D. Bai and J. Silverstein, Spectral analysis of large dimensional random matrices, Mathematics Monograph Series **2**, Science Press, Beijing 2006.
- [5] Z. D. Bai and Y. Q. Yin, Necessary and Sufficient Conditions for Almost Sure Convergence of the Largest Eigenvalue of a Wigner Matrix, *Ann. Probab.* **16** (1988), 1729-1741.
- [6] G. Bennett, *Probability Inequalities for the Sum of Independent Random Variables*, Journal of the American Statistical Association **57** (1962), 33-45.
- [7] O. Costin and J. Lebowitz, Gaussian fluctuations in random matrices, *Phys. Rev. Lett.* **75** (1) (1995) 69-72.
- [8] S. Dallaporta, V. Vu, A Note on the Central Limit Theorem for the Eigenvalue Counting Function of Wigner Matrices, *arXiv:1101.2553*
- [9] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Spectral Statistics of Erdős-Rényi Graphs I: Local Semicircle Law, *arXiv:1103.1919*.
- [10] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, Spectral statistics of Erdős-Rényi graphs II: eigenvalue spacing and the extreme eigenvalues, *arXiv:1103.3869*.
- [11] L. Erdős, B. Schlein and H.-T. Yau, Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices. *Ann. Probab.* **37** (2009), 815-852.
- [12] L. Erdős, B. Schlein and H.-T. Yau, Local semicircle law and complete delocalization for Wigner random matrices, *Comm. Math. Phys.* **287** (2009), no. 2, 641-655.
- [13] L. Erdős, B. Schlein and H.-T. Yau, Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Notices* **2010** (2010), 436-479.
- [14] L. Erdős, B. Schlein and H.-T. Yau, Universality of Random Matrices and Local Relaxation Flow, *arXiv:0907.5605*
- [15] L. Erdős, B. Schlein, H.-T. Yau and J. Yin, The local relaxation flow approach to universality of the local statistics for random matrices. *arXiv:0911.3687*
- [16] L. Erdős, H.-T. Yau, H.-T., and J. Yin, Bulk universality for generalized Wigner matrices. *arXiv:1001.3453*
- [17] L. Erdős, H.-T. Yau, and J. Yin, Rigidity of Eigenvalues of Generalized Wigner Matrices, *Advances of Mathematics*, to appear
- [18] P. Forrester, E. Rains, Interrelationships between orthogonal, unitary and symplectic matrix ensembles, Random matrix models and their applications, 171207, Math. Sci. Res. Inst. Publ., 40, Cambridge Univ. Press, Cambridge, 2001.
- [19] J. Gustavsson, Gaussian fluctuations of eigenvalues in the GUE, *Ann. Inst. H. Poincaré Probab. Statist.* **41** (2005), no. 2, 151-178.
- [20] D.L. Hanson, F. T. Wright, A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Math. Stat.* **42** (1971), no.3, 1079-1083.
- [21] S. Janson, T. Luczak, and A. Rucinski, Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000
- [22] K. Johansson, Universality of the local spacing distribution in certain ensembles of Hermitian Wigner matrices, *Comm. Math. Phys.* **215** (2001), no. 3, 683-705.
- [23] A. Knowles, J. Yin, The isotropic semicircle law and deformation of Wigner matrices, *arXiv:1110.6449*.

- [24] M. Ledoux, *A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices*, Séminaire de Probabilités XXXVII, 360369, Lecture Notes in Math., 1832, Springer, Berlin, 2003.
- [25] M. Ledoux, *Deviation inequalities on largest eigenvalues*, Geometric aspects of functional analysis, 167219, Lecture Notes in Math., 1910, Springer, Berlin, 2007.
- [26] M. Ledoux, B. Rider, *Small deviations for beta ensembles*, Electron. J. Probab. **15** (2010), no. 41, 1319-1343.
- [27] J. W. Lindeberg, Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung, *Math. Z.* **15** (1922), 211–225.
- [28] M. Meckes, Concentration of norms and eigenvalues of random matrices, *J. Funct. Anal.* **211** (2004), no. 2, 508-524.
- [29] S. O'Rourke, Gaussian fluctuations of eigenvalues in Wigner random matrices, *J. Stat. Phys.* **138** (2010), no. 6, 1045-1066.
- [30] T. Tao and V. Vu, Random matrices: Universality of the local eigenvalue statistics, *Acta Math.* **206** (2011), no. 1, 127–204.
- [31] T. Tao and V. Vu, Random matrices: universality of local eigenvalue statistics up to the edge, *Comm. Math. Phys.* **298** (2010), no. 2, 549–572.
- [32] T. Tao and V. Vu, Random matrices: Universality for the determinant of Wigner matrices, preprint, <http://arxiv.org/pdf/1111.6300.pdf>.
- [33] T. Tao and V. Vu, Random matrices: The universality phenomenon for Wigner ensembles, preprint.
- [34] C. Tracy and H. Widom, On orthogonal and symplectic matrix ensembles, *Commun. Math. Phys.* **177** (1996) 727–754.

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES CA 90095-1555

E-mail address: `tao@math.ucla.edu`

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN, CT

E-mail address: `van.vu@yale.edu`